DOMESTIC-ANIMAL GENOMICS: DECIPHERING THE GENETICS OF COMPLEX TRAITS

Leif Andersson[*‡] *and Michel Georges*[§]

One of the 'grand challenges' in modern biology is to understand the genetic basis of phenotypic diversity within and among species. Thousands of years of selective breeding of domestic animals has created a diversity of phenotypes among breeds that is only matched by that observed among species in nature. Domestic animals therefore constitute a unique resource for understanding the genetic basis of phenotypic variation. When the genome sequences of domestic animals become available the identification of the mutations that underlie the transformation from a wild to a domestic species will be a realistic and important target.

[*]Department of Medical Biochemistry and Microbiology, Uppsala University, Box 597, SE-751 24 Uppsala, Sweden.
[‡]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala Biomedical Centre, Box 597, SE-751 24 Uppsala, Sweden.
[§]Department of Genetics, Faculty of Veterinary Medicine, University of Liege (B43), 20, bd. de Colonster, 4000 Liege, Belgium.
Correspondence to L.A.
e-mail: leif.andersson@imbim.uu.se
doi:10.1038/nrg1294

During his voyage around the world, Charles Darwin's observations of the wealth of phenotypic diversity in nature inspired him to develop the theory of evolution by means of natural selection. However, this voyage took place more than 20 years before he published his seminal book *On the Origin of Species by Means of Natural Selection*[1]: in the meantime Darwin collected data to support his theory. The selective breeding of farm animals provided a large amount of these data. The phenotypic changes that were seen in farm animals that were subjected to selection essentially provided a 'proof-of-principle' for his theory. In fact, Darwin himself carried out breeding experiments with doves and he subsequently published two volumes on *The Variation in Animals and Plants under Domestication*[2].

Despite this early emphasis on the evolution of phenotypic variation among domestic-animal populations, these species were rapidly superseded as the models of choice after the rediscovery of Mendelian genetics. Practitioners of this new, largely laboratory-based, discipline favoured cheaper and easier-to-breed organisms with shorter generation times, such as the mouse and *Drosophila melanogaster*. Nonetheless, in terms of dissecting the genetic basis of phenotypic diversity, domestic animals have some notable advantages when compared with model organisms. No other animals have had their phenotypes monitored as closely as the principal domestic species. Moreover, thousands of years of selective breeding of these species has led to marked phenotypic changes and genetic adaptation to various environmental conditions. So, populations of domestic animals have a rich collection of mutations that affect phenotypic traits. Some of these traits, such as coat colour, have a simple monogenic basis, but most, such as growth, fertility and behaviour, are complex multifactorial characters.

The advantages of domestic animals will become increasingly important as we move into the post-genomic era. Despite the fact that we now know the complete or near-complete genome sequences of several organisms, our knowledge of the genes that underlie phenotypic differences within and among species is rudimentary. For example, human geneticists have had remarkable success in identifying the genes and mutations that underlie disorders with a simple monogenic inheritance, but the identification of genes that underlie disorders or traits with a complex genetic basis has proved difficult despite considerable efforts[3,4]. Similarly, our knowledge of the genetic basis of phenotypic variation among species is sketchy at best. Indeed, to "Understand evolutionary variation across species and

Box 1 | **Comparative genomics, phenotypic variation and domestic animals**

Comparative genomics is an important approach for dissecting the genetic basis of phenotypic variation. For example, there is considerable interest in comparing human and chimpanzee genomes to identify the genes that underlie the phenotypic traits that make us human[53–55]. However, comparing genomic sequences among species does not allow functionally important substitutions to be easily identified, unless a mutation occurs in the coding sequence of a gene with a well-known function. The analysis of phenotypic extremes within species might provide a complementary approach to among-species comparative genomics. Charles Darwin used domestic-animal species as model organisms and they might once again provide a model system for understanding phenotypic evolution. Different breeds of domestic animals are in many cases as phenotypically diverse as separate species (see figure). However, the recent origin of these breeds from their wild ancestors (~10,000 years before present) makes it possible to make specific crosses and use segregation analysis to map the genes that underlie phenotypic traits. Moreover, owing to their recent origin, the sequence divergence among breeds of domestic animals is negligible, which makes it easier to identify the causative mutations that underlie phenotypic differences. By contrast, humans and chimpanzees diverged from a common ancestor millions of years before present and the average sequence divergence between the two species is ~1.2% (REF. 55). This corresponds to ~40,000,000 nucleotide substitutions plus numerous insertion and/or deletion differences. So, functionally important substitutions between humans and chimpanzees constitute the tip of an iceberg of changes that are selectively neutral.

Nonetheless among-species comparative genomics will be crucial for pinpointing functionally important, non-coding sequences. The comparison of the human and mouse genome sequences showed that ~5% of the mammalian genome is evolutionary conserved[56]. However, only ~1.5% of these genomes encode proteins, which indicates that the remaining evolutionarily conserved fraction of these genomes (~3.5%) is non-coding. It is clear that the comparison of genome sequences from numerous species markedly improves the chance of recognizing evolutionarily conserved sequences[57,58]. A striking illustration of this concept is provided by the recent identification of the causative mutation underlying the insulin-like growth factor 2 (*IGF2*) quantitative trait locus (QTL) in pigs (BOX 3). The mutation occurs in the middle of an intron and was identified by genetic analysis. A bioinformatic analysis of sequences from eight mammalian species showed that the 100 base pairs (bp) flanking the mutation have ~85% sequence identity among distantly related mammals, as high as for most coding sequences, and the mutation is part of a 16-bp segment that has 100% identity among all eight species[7]. Consequently, sequence variation that is found in such evolutionary conserved regions can be used as primary candidates for phenotypic differences. The photograph of the boy with a female gorilla was provided by L. A.

the mechanisms underlying it" has been named as one of the 'grand challenges' of future genomics research[5].

Comparative genomics will have a big role in addressing this challenge. Comparative analyses of the genomes of different domestic breeds might prove to be one of the most efficient ways of dissecting the genetic basis of phenotypic variation. The large phenotypic differences and the limited neutral genetic variation among breeds make them ideal candidates for study (BOX 1). The identification of the mutations that underlie the variation of several interesting monogenic phenotypic traits in domestic animals (TABLE 1), as well as some mutations that underlie complex traits[6–10], has already illustrated the potential of domestic animals for uncovering the genes that underlie phenotypic diversity. It is also worthwhile noting that many quantitative trait loci (QTLs) that affect a broad range of phenotypes — including growth, body composition and fertility — have already been mapped with high confidence in the different livestock species and are awaiting further characterization (for example, see REF. 11).

However, the lack of genomic resources in domestic animals, as compared with species such as human and mouse, has hampered progress in gene mapping and identification. This situation will change markedly in the near future with the completion of draft genome sequences for key domestic animals. The chicken and dog sequences are already underway and the cow will follow soon after. Animal geneticists will then have a near complete list of all coding sequences, their chromosomal location, numerous genetic markers and the possibility to generate gene arrays for highly informative expression analyses. So, it will soon be possible to exploit the full potential of farm-animal genomics. Here, with this potential in mind, we provide an overview of domestic-animal genomics and the potential boost it will offer to our future understanding of complex traits. First, we discuss the challenge that is presented by the genetic analysis of multifactorial traits and the way that this challenge has been addressed in domestic animals. We then summarize the present status of the diverse array of domestic-animal genome projects that are underway. Finally, we outline how the completion of these genome projects will facilitate complex-trait analysis in the future.

The challenge of multifactorial traits

Most biological traits and all common diseases in humans have a multifactorial (or complex) inheritance, which indicates that they are influenced by numerous genes and environmental factors. A chromosomal

Table 1 | **Monogenic trait loci for which the causative mutation has been identified**

Species	Trait	Gene	Reference
Cattle	Muscle hypertrophy	*MST*	72–75
	Coat colour	*MC1R*	76
	White coat colour	*KITLG*	23
	Fish odour in milk	*FMO3*	77
Chicken	Albinism	*TYR*	78
	Plumage colour	*MC1R*	79
	Dominant white plumage colour	*PMEL17*	80
Dog	Narcolepsy	*HCRTR2*	81
	Coat colour	*MC1R*	82
Goat	Lack of horns, intersexuality	Non-coding region*	83
Horse	Coat colour	*MC1R*	84
		ASIP	85
		MATP	86
	White colour, megacolon	*EDNRB*	87–89
Pig	Malignant hyperthermia	*RYR1*	21
	Dominant white colour, haematopoiesis	*KIT*	22,90,91
	Hypercholesterolaemia	*LDLR*	92
	Coat colour	*MC1R*	93,94
	Intestinal *Escherichia coli* adherence	*FUT1*	95
	Glycogen content in skeletal muscle	*PRKAG3*	24
Sheep	Fertility, ovulation rate	*BMP15*	96
		BMPR1B	97
	Muscle hypertrophy	Regulatory mutation*	98,99

*This is apparently a regulatory mutation that affects the expression of one or more genes in the chromosomal region to which it maps.

region that contains one or more genes that influence a multifactorial trait is known as a QTL[12,13]. The use of segregation analysis in informative families or experimental crosses to map QTLs is well established[11,14,15]. The power of such analyses to detect and map QTLs depends on how large a fraction of the phenotypic variation is explained by a given locus and the size of the segregating population.

The principal challenge with multifactorial traits lies not in detecting QTLs, but in unravelling the genes that underlie them. Despite large efforts to identify the genes that affect multifactorial traits, in particular those that are involved in common human diseases, there are few success stories[3,4]. The identification of genes and mutations that underlie QTLs is problematic for several reasons. First, it remains difficult to determine the exact chromosomal location of a QTL. As for monogenic Mendelian traits, the marker and crossover density in the region of interest limits the mapping precision. The fuzzy 'detectance' of QTLs, that is, the probability of a QTL genotype given the phenotype, complicates matters even more. The lack of a direct relationship between genotype and phenotype, as exists for monogenic traits, prohibits the unambiguous identification of recombinant individuals that is required for high-resolution mapping. This is due to the fact that individual QTLs only account for part of the phenotypic variance, the rest being due to environmental factors as well as other QTLs. The situation is particularly tricky in the case of numerous loosely linked QTLs, which seem to account for a significant proportion of the largest QTL effects that are detected in livestock (M.G., unpublished observations). Although COMPOSITE INTERVAL MAPPING AND MULTIPLE QTL MAPPING might help to unravel such situations in experimental crosses between inbred lines[14], the development of a suitable statistical framework to address such situations in outbred designs is only in its infancy[100]. EPISTATIC interactions might also add to the challenge of dissecting the genetic basis of complex traits (BOX 2). Certainly, data from model organisms and from coat-colour inheritance in mammals indicate that epistasis between QTLs might be an important factor for consideration. However, studies of epistasis among QTLs in vertebrates have been rare[16–18]. For these reasons, QTLs are often mapped to chromosomal regions that are over 20 centiMorgan (cM) long (~20 megabase pairs (Mb)) and that might contain several hundred genes.

Second, most QTLs have a mild phenotypic effect, so the mutations that cause them are difficult to distinguish from neutral polymorphisms. By contrast, mutations that cause monogenic disorders generally knock out gene expression or lead to an altered protein function. Another factor that complicates the identification of QTL mutations is that it is likely that a good proportion of these are regulatory mutations. Our ability to spot and evaluate functionally important mutations in non-coding regions is still poorly developed. For example, the identification and annotation of regulatory regions in sequenced genomes is still very rudimentary compared with that of coding sequences, although our ability to identify important regulatory elements through sequence comparison will improve as more

COMPOSITE INTERVAL MAPPING AND MULTIPLE QTL MAPPING
Methods that increase quantitative trait locus (QTL) mapping resolution in a chromosome interval of interest, by accounting for genetic background noise due to segregation at other QTLs by means of the inclusion of multiple markers as cofactors in the statistical model.

EPISTASIS
The phenotypic expression of genotypes at one locus depends on the genotype at another locus or other loci.

EFFECTIVE POPULATION SIZE
The number of individuals in a theoretically ideal population that are subject to the same amount of genetic drift as the actual population

HETEROZYGOSITY
The frequency of heterozygotes at a locus.

LINKAGE DISEQUILIBRIUM
The non-random association of alleles at different loci.

MALIGNANT HYPERTHERMIA AND HALOTHANE SENSITIVITY
A disorder in which uncontrolled muscle contractions can cause lethal overheating. In pigs, this pathological condition might be induced by stress or exposure to halothane anesthesia. Susceptibility to malignant hyperthermia in pigs and in some human families is caused by mutations in the *RYR1* gene, which encodes ryanodine receptor 1.

MÜLLERIAN DUCTS
The structures from which the vagina, cervix, uterus and oviducts derive in the female embryo.

EPIGENETIC INHERITANCE
Inheritance of a molecular modification of DNA (methylation or chromatin structure) that affects gene expression.

mammalian genomes are sequenced (BOX 1). The situation might be even more complicated if QTL effects reflect the combined action of clusters of tightly linked mutations[19], or if epigenetic inheritance contributes to quantitative genetic variation (BOX 2).

Genetic dissection of complex traits

So, it is clear that dissecting the genetic basis of complex traits presents a significant analytical challenge, and for this reason the systems in which we have more power to detect QTLs and the mutations that underlie them are of particular interest. Domestic animals are one such system. For several reasons, the power to detect the mutations in QTLs that underlie variation in complex traits is much better in domestic animals than in human families. First, there is less genetic heterogeneity within breeds owing to the limited EFFECTIVE POPULATION SIZE compared with large outbred human populations. This means that the number of segregating QTLs and the number of alleles at each locus that affect a certain trait is expected to be lower within populations of domestic animals. Second, large family sizes in domestic animals (hundreds or thousands of progeny in some species) make it possible to deduce the QTL genotype of the parents with confidence by using progeny testing — that is, HETEROZYGOSITY can be deduced on the basis of the presence of QTL segregation, and homozygosity can be inferred by the lack of segregation if the family size is sufficiently large in relation to the expected size of the QTL effect[13,20]. Nonetheless, there are still relatively few examples for which the mutations that underlie mapped QTLs have been identified in domestic animals. These few examples have been identified either because a gene that causes a monogenic trait has pleiotropic effects on several complex traits, or by adopting a positional candidate approach combined with LINKAGE-DISEQUILIBRIUM analysis.

Genes with pleiotropic effects on complex traits. The identification of genes that cause monogenic traits is straightforward. The direct relationship between genotype and phenotype allows the gene responsible to be mapped with high resolution. However, a gene that has a large effect on a monogenic trait can have minor effects on other complex traits. In effect, such a gene is behaving as a QTL for the complex traits that it influences. This is well illustrated by some examples from domestic animals. For instance, a missense mutation in the pig *RYR1* gene, which encodes a calcium channel that is expressed in skeletal muscle, causes MALIGNANT HYPERTHERMIA AND HALOTHANE SENSITIVITY in the homozygous condition[21]. However, this mutation is also associated with high lean-muscle content (that is, more muscle and less fat). Therefore, strong selection for leaner pigs between 1960 and 1990 markedly increased the frequency of this mutation during that period. As a result malignant hyperthermia in commercial pig lines was a major practical problem in commercial pig production until a DNA test for this mutation became available.

Other examples come from genes that influence the colour of domestic animals. The combined effect of a gene duplication and a splice mutation in the *KIT* gene causes dominant white-coat-colour in pigs[22]. However, these mutations have pleiotropic effects on haematopoieses and the locus behaves as a QTL that influences the number of red and white blood cells. A parallel example is a missense mutation in the KIT ligand gene (*KITLG*) that causes roan/white colour in cattle and is associated with developmental anomalies of MÜLLERIAN DUCTS (also known as white heifer disease) with complex, multifactorial inheritance[23].

The positional candidate approach. In the positional candidate strategy linkage analysis is used to map the locus to a specific chromosomal region. This region is subsequently scrutinized for candidate genes that might influence the trait being studied. The next step is then to search for causative mutations in the candidate gene. This approach is the most common strategy for dissecting monogenic traits in mammals as there are complete genome sequences for several species and the functional characterization of genes is continuously improving. However, the generally poor resolution of initial QTL mapping means that this approach is more difficult to apply to multifactorial traits. Specifically, the region that a QTL is mapped to might contain too many plausible candidate genes and even several poorly characterized genes that cannot be excluded as candidates. Occasionally, however, positional candidate cloning can be a quick way of identifying a causative gene that can be confirmed by further genetic data or functional assays.

An example of the successful use of this approach is the identification of allelic variants at the *PRKAG3* locus that encodes a subunit of AMP-activated protein kinase (AMPK), an enzyme that has a key role in the metabolic

Box 2 | Epistasis, epigenetic inheritance and quantitative trait loci

There is good evidence that epistasis between quantitative trait loci (QTLs) has an important effect on multifactorial traits. In particular, the significance of epistatic interaction among the genes that control coat colour in mammals is well established[59], and data from experimental organisms highlight the importance of epistasis[60]. So, some QTLs will only have an effect on certain genetic backgrounds. However, the significance of epistatic interactions has not yet been extensively studied in non-experimental organisms, partly owing to an expected lack of statistical power in most studies and partly because of the lack of computer software that can handle the demanding statistical analysis. However, recent studies indicate that epistatic interaction contributes significantly to quantitative variation[16–18].

Does EPIGENETIC INHERITANCE contribute to the complex genetic background for multifactorial traits? Although the general model is that epigenetic imprints are erased during the development of a new individual[61], there is evidence that epigenetic imprints in the form of DNA-methylation patterns or chromatin configuration can be transmitted from parent to offspring and could influence the phenotype of the offspring. One example is the expression of coat colour in mice that have inherited the *Viable yellow* allele at the *Agouti* locus from their dams[62]. In this case the incomplete erasure of an epigenetic modification at a retrotransposon, inserted upstream of the *Agouti* gene, when the *Viable yellow* allele was transmitted through the maternal germline, affects the phenotype. Evidence for epigenetic inheritance is also well documented in yeast[63], *Drosophila melanogaster*[64,65] and plants[66–68]. This leads to the unorthodox view that a portion of the inherited variation in multifactorial traits can not be explained by differences in the nucleotide sequence itself, but in the degree of methylation or in the chromatin configuration.

ADVANCED INTERCROSS LINES
The subsequent generations (F_3, F_4 and so on) of an intercross, which are maintained to allow the high-resolution mapping of quantitative trait loci.

HAPLOTYPE
A combination of alleles at different loci that is transmitted together from one generation to the next.

regulation of eukaryotic cells. Pigs with the so-called RN phenotype have a high glycogen content in skeletal muscle and a high lean-meat content, but the quality of the meat is not as good for processing as that from normal pigs. A pure positional cloning approach was used to identify the causative mutation as a missense substitution (R225Q) in *PRKAG3* (REF 24). This mutation might be considered as a QTL allele, but the phenotypic difference between genotypes is so large that this trait is effectively monogenic. However, Ciobanu *et al.*[25] subsequently found a QTL peak for several meat-quality traits including glycogen content that mapped to the same region as *PRKAG3*. The R225Q mutation was not involved, but three other missense mutations in this gene were potential candidates. Further studies in several commercial pig lines have confirmed that at least of one of these mutations — at a residue adjacent to the R225Q mutation (V224I) — influences glycogen content and meat quality[25–27]. Interestingly, it has recently been shown that these two residues are located in a part of AMPK that is directly involved in the binding of AMP: a key step in the allosteric activation of the enzyme[28].

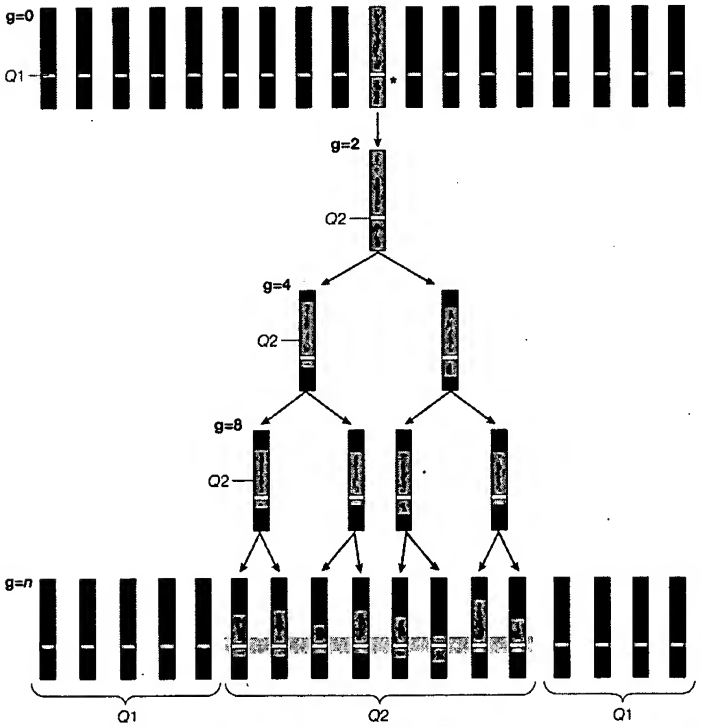


Figure 1 | **Identical-by-descent mapping.** Assume that the quantitative trait locus (QTL) allele *Q2* originates by mutation from allele *Q1* at generation 0. There will be a complete linkage disequilibrium between *Q2* and alleles at all other loci in the first gamete carrying *Q2*. This linkage disequilibrium will then gradually decay through each generation owing to recombination, but linkage disequilibrium will persist for closely linked loci. At generation *n* a sample of chromosomes are collected and classified (*Q1* or *Q2*) by segregation analysis. Genetic markers and sequence analysis are then used to define the minimum haplotype that is shared identical by descent among animals carrying *Q2* (indicated by the yellow bar).

Identical-by-descent mapping. In model organisms, further breeding experiments or ADVANCED INTERCROSS LINES can be used to refine the map position of QTLs[29]. However, with the exception of the chicken, the application of such an approach to domestic animals is expensive. A promising alternative approach is to combine linkage and linkage-disequilibrium analysis[30–33]. The basic principle of this approach is outlined in FIGURE 1. Assume that a QTL allele *Q1* mutates to *Q2* at generation 0. In the first generation there is complete linkage disequilibrium between *Q2* and the alleles at all other polymorphic loci on the same chromosome. In each subsequent generation, recombination gradually reduces the size of the block of linkage disequilibrium that surrounds the QTL. The key to narrowing down the location of the QTL is to use linkage analysis to deduce the QTL genotype, and then to use a dense set of genetic markers to determine the minimum HAPLOTYPE that is shared identical by descent (IBD) among the animals that carry the *Q2* allele.

The recent and strong selection that domestic animals have been subject to make the IBD-mapping approach useful in these species. Strong directional selection in domestic animals has led to selective sweeps in which alleles at loci that underlie selected traits have increased markedly in their frequency. This process leads to a loss of heterozygosity in the flanking region owing to 'hitch-hiking'[34] (FIG. 2). The size of the region that shows a hitch-hiking effect will depend on how quickly the favourable haplotype becomes fixed (homozygous) and the recombination rate in the interval. After a selective sweep the occurrence of new mutations slowly restores the heterozygosity. However, the short evolutionary history of animal domestication indicates that the genomic footprints of major selective sweeps should largely remain.

The IBD approach was used in cattle to position a QTL for twinning rate to an interval of less than 1 cM[35]. It was also used to identify putative causative mutations in the diacylglycerol acyltransferase (*DGAT*) and growth hormone receptor (*GHR*) genes, which underlie two major QTLs for milk-production traits on cattle chromosome 14 and 20, respectively[6,8,10]. Subsequent genetic and functional studies have now provided strong evidence that a missense mutation in *DGAT* (K232A) is the mutation that underlies a QTL for milk yield and composition[9]. The insulin-like growth factor 2 (*IGF2*) QTL in pigs is another prime example of how this approach has allowed extraordinary resolution in QTL mapping, down to a single nucleotide substitution in a non-coding region[7,20] (BOX 3).

The *IGF2* story illustrates some of the advantages of using domestic animals for QTL studies. The genetic evidence pinpointing the mutation that underlies the *IGF2* QTL was obtained because the ancestral haplotype, which only differed at the quantitative trait nucleotide (QTN), was available. There is a better chance of the ancestral haplotype being available in domestic species than for most other species. This is partly because selective sweeps in domestic animals have often occurred within a fairly short period of time,

which increases the chance that the ancestral haplotype is still present in some populations. Moreover, in principle, animal geneticists have access to the world population of domestic animals, so the entire diversity of haplotypes that is present in these species can be characterized. By contrast, laboratory strains of most experimental organisms represent only a tiny fraction of the genetic diversity of these species[36,37].

The mapping of the *IGF2* QTN also illustrates the advantage that the QTL-mapping approach to finding genes that control multifactorial traits might have compared with mutagenesis-screening programmes (which are often considered to be a more productive way of approaching this task[38]). The *IGF2* QTN only has a limited effect on muscle mass in the pig (a 3–4% increase), so it would be extremely difficult to identify the causal mutation in this case with any other genetic screening method available at present. In particular, this limited effect would prevent its identification in a high-throughput mutagenesis–screening programme.

Domestic-animal genomics

The basic tools for genome research have been established for all principal domestic species: hundreds of microsatellite markers; low-resolution linkage and physical maps; and large-insert genomic libraries (see the online links box). However, despite the notable successes that are discussed above, positional cloning of trait loci in domestic animals has been hampered by the absence of high-resolution linkage maps (several markers per cM), comprehensive collections of expressed sequence tags (ESTs), whole-genome bacterial artifical chromosome (BAC) contigs and whole-genome sequences. Positional cloning of monogenic trait loci and QTLs have therefore relied on comparative mapping using primarily the human and mouse maps, and more recently genome sequences[13]. This is a laborious approach as it is often necessary to design numerous primer pairs for PCR that are based on conserved sequences from other species, and to generate local BAC contigs for each region of interest. Furthermore, there is always the risk of a minor chromosomal rearrangement between the target and reference species that might slow down progress. This approach is particularly cumbersome in the chicken owing to the large evolutionary distance between birds and mammals (~300,000,000 years).

However, the development of extra genome resources and the imminent completion of draft genome sequences for several of the principal domestic animals should soon remove the bottlenecks that are hampering the positional cloning of QTLs in domestic animals (see TABLE 2 and below). Efforts are now underway to generate high-quality draft (HQD) genome sequences for chicken, dog and cattle. These will not be finished genome sequences, so there will be sequence gaps and there will be errors in the assembly. However, it will be relatively straightforward for researchers to generate finished sequences for their regions of interest using the HQD sequence combined with the finished sequence from other vertebrates.

Chicken. Chicken will be the first domestic animal to have its genome sequenced to near completion. At present, the Washington University School of Medicine Genome Sequencing Center is completing a HQD genome sequence that is based on 6.6× WHOLE-GENOME SHOT-GUN (WGS) sequence and end-sequencing of large-insert clones combined with the generation of a BAC FINGERPRINT MAP (for chicken genome projects, see online links box). The sequence is based on genomic DNA from a single red junglefowl female. Sequence reads providing ~6× coverage have already been deposited in GenBank's trace archive (see online links box) and a draft genome assembly will be released during February 2004 (TABLE 2). The assembly of the genome sequence will be facilitated by the small genome size (~1.1 gigabase pairs (Gb)) and the low frequency of repetitive sequences compared with mammalian genomes. The sequence that is generated using the red junglefowl (which is the wild ancestor of the domestic chicken) will be complemented with a sequencing effort by the Beijing Genomics Institute (see online links box) that is expected to generate ~1× coverage using genomic DNA from three breeds of domestic chickens (TABLE 2).

Dog. A partial genome sequence (1.5× coverage) that is based on the sequence of a poodle has already been published[39]. The bioinformatic analysis of these data identifed fragments representing putative orthologues of ~75% of all annotated human genes, and more than 4% of the non-coding sequence was found to be conserved between dog, human and mouse. The average sequence identity for orthologous genes was higher between dog and human than between human and mouse, despite the evidence that the dog is the phylogenetic outgroup of these three species. This result is explained by a higher substitution rate in the rodent lineage. This survey sequence is a valuable resource for dog genomics. As an example, almost 1,000,000 putative single nucleotide polymorphisms (SNPs) and 150,000

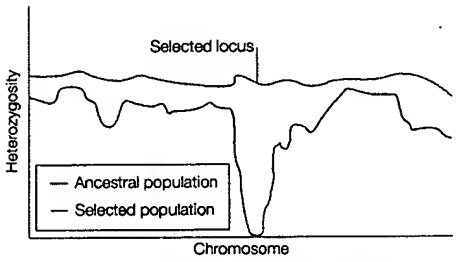


Figure 2 | **Loss of heterozygosity owing to a selective sweep of a favourable mutation.** In an outbred population the heterozygosity varies along a chromosome according to the local mutation rate, previous selection history and genetic drift. Strong directional selection in domestic animals (and in other species) is expected to cause selective sweeps in which a favourable allele replaces other alleles. This leads to homozygosity at the selected locus and also at flanking loci owing to 'hitch-hiking'[34]. This characteristic pattern means that dense genome scans can show regions of the genome that have gone through selective sweeps.

WHOLE-GENOME SHOT-GUN
The random generation of short DNA-sequence reads from the whole genome.

FINGERPRINT MAP
A map of a clone or a genome that is based on the pattern of fragments that are generated by restriction enzyme digestions.

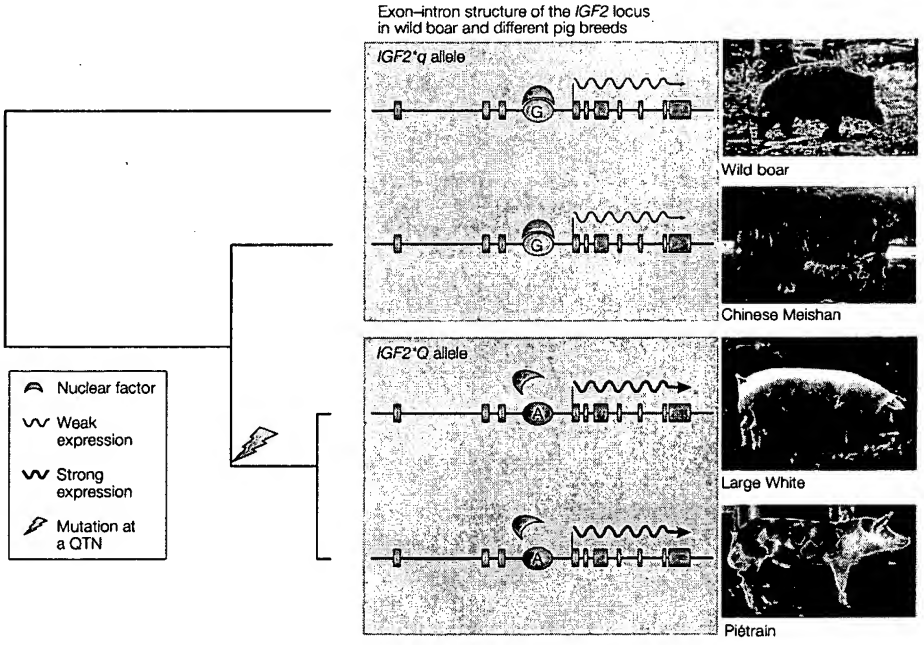
polymorphic microsatellites were identified as heterozygous positions in the single dog that is being sequenced.

A HQD genome sequence (~6.5× coverage) of another breed of dog (a boxer) will be completed by Spring 2004 (TABLE 2). Moreover, a large-scale SNP discovery project to identify SNPs that are common to many breeds is also underway. The HQD will be an important leap forward for dog genomics and will facilitate the identification of the causative mutations for some of the plethora of monogenic disorders that are found in dogs[40]. The recent identification of a mutation that causes renal cystadenocarcinoma and nodular dermatofibrosis in the German Shepherd dog illustrates the potential of this approach[41]. Furthermore, the rich and interesting phenotypic diversity in the morphology and behaviour in dogs (for example, see BOX 1) will be

Box 3 | The *IGF2* quantitative trait locus in pigs

Insulin-like growth factor 2 (*IGF2*) was first identified as a paternally expressed quantitative trait locus (QTL) in INTERCROSSES between the European wild boar and Large White domestic pigs; and between Large White and Piétrain pigs[69,70]. In the wild-boar intercross, the QTL allele from the domestic pig was associated with high muscularity, less backfat and a larger heart. Sequence analysis showed that the *IGF2* haplotypes in the Swedish Large White and Piétrain pigs were identical by descent, whereas the Belgian Large White and wild-boar haplotypes were similar, which indicated the presence of two alleles that are denoted *Q* and *q* for high and low muscle growth, respectively[7]. Another intriguing observation was the large sequence divergence (~1%) between the two haplotypes. This led to the suspicion that the two haplotypes might have an Asian and European origin, in line with previous finding that some European breeds, including the Large White, are hybrids of Asian and European domestic pigs[48,71]. Sequence analysis of *IGF2* haplotypes segregating in an intercross between Chinese Meishan and Large White pigs confirmed this hypothesis. The Meishan allele, which was functionally equivalent to *IGF2*q*, was nearly identical to *IGF2*Q* at the sequence level. These data provided conclusive evidence that the causative mutation for *IGF2*Q* was a G-to-A substitution at nucleotide 3,072 in intron 3. *IGF2* was identified as a positional candidate gene, but the quantitative trait nucleotide (QTN)[12] was identified by pure genetics: linkage analysis to deduce QTL genotypes combined with an analysis of the minimum shared haplotype.

Functional studies showed a plausible mechanism for the QTL effect (see figure). First, the mutation does not affect the imprinting or METHYLATION STATUS of the QTN region and the region is undermethylated in skeletal muscle. Second, the wild-type sequence binds a nuclear factor and this interaction is abrogated by the mutation and by methylation. Third, transfection analysis indicated that the wild-type sequence functions as a silencer element, whereas the mutant sequence is a significantly weaker silencer. Finally, expression analysis showed an approximately threefold upregulation of *IGF2* expression in postnatal skeletal and cardiac muscle but not in prenatal muscle or in liver. The result is consistent with phenotypic data showing that *IGF2*Q* are associated with high muscle growth and a larger heart, but has no effect on birth weight or the size of the liver. The *IGF2* QTL is truly adaptive from a pig production point of view as it does not affect birth weight but supports muscle growth after birth. The photographs of the wild boar, Meishan, Piétrain and Large White pigs were provided by B. Kristiansson, Quality Genetics AB, J.-M. Beduin and the Roslin Institute, respectively.



INTERCROSSES
Crosses between different populations.

METHYLATION STATUS
The degree of methylation of the cytosine base in DNA.

Table 2 | **Present status of efforts to determine the genome sequences of domestic animals**

Species	Genome size	Number of ESTs*	Institute	Coverage	Time schedule
Chicken	~1.1 Gb	451,655	WashU	6.6× WGS	February 2004‡
			Beijing	1× WGS	Spring 2004§
Dog	~2.8 Gb	27,010	TIGR	1.5× WGS	Published (see REF. 39)
			Broad	6.5× WGS	Spring 2004‖
Cattle	~3 Gb	331,140	Baylor	7× WGS	2005¶
Pig	~2.7 Gb	240,001	Beijing	~1× WGS	2004#
Horse	~3 Gb	15,240	–	–	–
Cat	~3 Gb	228	–	–	–
Sheep	~3 Gb	6,748	–	–	–

*GenBank, 9 January 2004. ‡Wesley Warren, personal communication. §Bin Liu, personal communication. ‖Kerstin Lindblad-Toh, personal communication. ¶George Weinstock, personal communication. #Merete Fredholm and Bin Liu, personal communication.
Baylor, Human Genome Sequencing Center at Baylor College of Human Medicine; Beijing, Beijing Genomics Institute; Broad, Broad Institute; EST, expressed sequence tag; Gb, gigabase pairs; TIGR, The Institute for Genomic Research, Rockville; WashU, The Genome Sequencing Center at Washington University School of Medicine; WGS, whole-genome shot-gun (See associated web sites in the online links box).

able to be studied in much more detail when the HQD sequence becomes available. However, the identification of the mutations that underlie interbreed diversity will be challenging as few attempts have been made to generate or collect informative intercross pedigrees that can be used for genetic studies.

Cattle. An effort to generate a HQD genome sequence from cattle has been initiated and the plan is to carry out BAC SKIM SEQUENCING (~1× coverage), 5× WGS reads from a single, partially inbred Hereford animal, and finally 1× WGS reads from animals representing different breeds to allow extensive SNP detection (TABLE 2). In addition, a BAC fingerprint map is now being generated for cattle (for cattle genome projects, see online links box). These will be important resources for the ongoing efforts to identify genes that influence milk- and meat-production traits, as well as the genes that influence susceptibility to infectious diseases in this species. The cattle genome sequence will also be a valuable resource for sheep and goat genomics owing to the close evolutionary relationship between these species (~20,000,000 years).

Pig. A Chinese and Danish collaboration has generated about 800,000 ESTs from several pig tissues and a partial genome sequence (~1× coverage). This sequence information will be released during 2004 (TABLE 2). Unfortunately, no funding has yet been secured for a HQD genome sequence and the National Human Genome Research Institute (NHGRI) has given medium priority to the sequencing of the pig genome (see online links box). Pig genomics will benefit to some extent from access to a HQD genome sequence for cattle, but these species diverged early during the evolution of the *Cetartiodactyla* lineage, about 60,000,000 years before present[42].

SNP detection. The WGS efforts for chicken, dog and cattle will primarily be based on a single individual as the assembly of the genome sequence is facilitated by a reduced genetic heterogeneity. The drawback with this approach is that the sequence data will be less informative for finding polymorphisms. In the chicken, this will be compensated by the generation of ~1× genome coverage from domestic chickens, which is expected to reveal millions of chicken SNPs. Similar efforts are underway in dogs and in cattle (see above; see also the USDA Meat Animal Research Center (MARC) genomic resources in online links box). There are also initiatives to detect numerous SNPs from available EST resources in chicken (see BBSRC ChickEST database in the online links box), cattle[43] (for cattle genome resources, see the online links box) and in pigs[44] (M. Fredholm, personal communication).

SKIM SEQUENCING
The partial random sequencing of a large-insert clone.

The future of complex-trait analysis

The generation of HQD genome sequences will allow the full potential of using domestic animals for deciphering the genetic basis of multifactorial traits to be realised. The access to the genome sequence will speed up QTL detection in several ways (FIG. 3). Low-resolution QTL mapping in commercial and experimental populations is already well established. High-resolution IBD mapping, which is vital for positional cloning of QTLs, will soon be facilitated by the access to numerous microsatellite and SNP markers, and possibly haplotype maps as have been developed for inbred strains of mice[45] and for humans[46]. The identification of candidate genes and mutations will be facilitated by the access to species-specific genome sequences, and the re-sequencing of QTL intervals will be an attractive approach if the achieved map resolution is sufficiently good.

The access to dense marker maps should open up the possibility for a new approach for QTL detection. If the marker density is sufficiently high then marker screenings using a limited number of animals (or pools of animals) representing different populations of domestic animals should detect the footprints of selective sweeps, such as the one observed for *IGF2* in pigs (FIG. 2). Several factors influence the expected size of haplotype blocks that are fixed by selective sweeps. A high recombination rate will reduce the length of haplotype blocks. If the effective population size is small, large haplotype blocks might be fixed owing to genetic drift. For example, inbred lines of laboratory mice have large haplotype blocks and

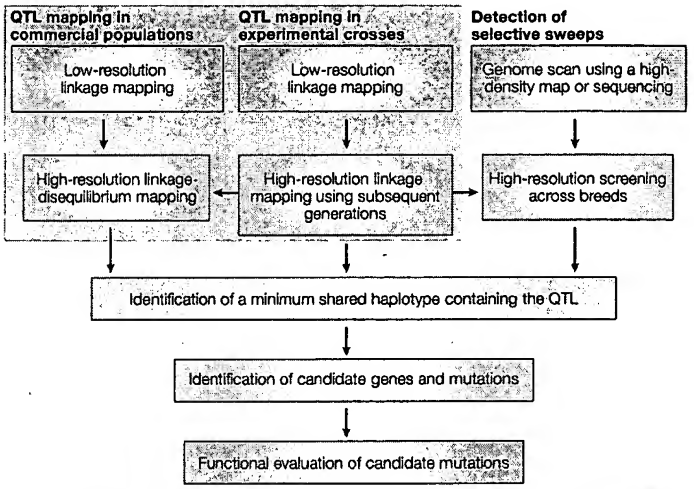


Figure 3 | **Approaches to mapping and positional cloning of QTLs in domestic animals.** The segregation of quantitative trait loci (QTLs) can be detected in family material from commercial populations or from experimental crosses. QTL mapping in intercrosses between divergent populations has an excellent power for QTL detection, owing to the high heterozygosity at QTLs in the F_1 generation. However, the resolution is rather poor as it is based on those recombination events that occur in the experimental pedigree. So, an initial low-resolution mapping in an intercross can be followed up by high-resolution linkage-disequilibrium mapping within a commercial population, if the QTL is segregating within such populations, or by the detection of the minimum shared haplotype representing a selective sweep. The identification of the causative mutation for the insulin-like growth factor 2 (*IGF2*) QTL in pigs is an excellent illustration of how this combined approach has been used (BOX 3).

most of these have been fixed by drift[45]. In certain dog breeds it can be difficult to distinguish haplotype blocks that have been fixed by selection from those that have been fixed by drift, as many breeds have been established with a limited number of founder animals or have gone through severe population bottlenecks. The size of haplotype blocks is strongly influenced by the selection intensity as it determines how quickly a haplotype carrying a favourable mutation reaches fixation. So, modern breeding programmes with intense selection are expected to cause fixation of larger haplotype blocks compared with the less intensive animal breeding that was carried out before the twentieth century. However, even QTL alleles with fairly large effects are not expected to be fixed in a few generations as they only explain, by definition, a fraction of the genetic variance for a given trait.

There is an inherent conflict concerning the optimal size of haplotype blocks to be used for the detection of QTLs. The larger a fixed haplotype block is the easier it is to detect, but it will then be more difficult to identify the causative gene and mutation. The general trend is that population-wide linkage disequilibrium in livestock extends over tens of cM rather than <1cM regions, as typically observed in humans[47]. However, it is often the case that there are several breeds of domestic animals that are selected for the same purpose and there is often some gene flow between similar breeds. So, the same IBD segment might be favoured by selection in different breeds. This is very well illustrated by the *IGF2* QTL in pigs[7]. The same IBD segment was identified in four different breeds of pig. The region of strong linkage disequilibrium within breeds extended over several-hundred kb, but the minimum IBD haplotype that was identified across breeds was only 20 kb. Furthermore, many domestic breeds have a mosaic structure of variation owing to recent or historical ADMIXTURE of divergent breeds. The INTROGRESSION of Asian domestic pigs into European domestic breeds during the eighteenth and nineteenth century is one good example of recent admixtures[48]. Similarly, some breeds of African cattle are the products of a progressive, male-driven admixture between the indigenous taurine breeds (*Bos taurus*) and immigrating zebu cattle (*Bos indicus*)[49]. In both of these examples admixture has occurred between populations that originate from different subspecies of the wild ancestors of the livestock species. This situation can in fact greatly facilitate the detection of selective sweeps owing to the sequence divergence between haplotypes that originate from different subspecies (BOX 3).

So, the marker density that is required to detect haplotype blocks that are fixed by selective sweeps will vary from population to population on the basis of its history, and from locus to locus depending on the local recombination rate and how quickly the favourable allele reached fixation. Suitable marker densities can be established empirically using test cases such as myostatin in cattle, *IGF2* in pigs, as well as coat-colour loci: a marker density of at least 10 markers per cM will be required for QTLs, which corresponds to ~30,000 markers for a genome-wide scan. Ideally WGS reads would be generated using representatives of different breeds to 0.1–1× coverage to uncover IBD regions within and across breeds, or even better to >2× coverage to uncover most variants. The cost of SNP genotyping or sequencing will, in the near future, limit the practical application of this IBD scanning approach. However, it should be an attractive approach for selected regions of the genome that contain important QTLs and the access to HQD genome sequences will make the approach feasible.

The IBD scanning method of domestic breeds and their wild ancestors might also uncover chromosomal regions that contain important mutations that have been crucial during domestication. Such regions are expected to affect behaviour, reproduction, production and possibly coat colour. Furthermore, the comparison of animals representing breeds that are selected for different purposes, such as dairy and beef cattle and egg- and meat-producing chickens, should uncover the loci that have responded to selective breeding as those showing low diversity within breeds but high diversity between divergently selected breeds. The proposed approach does not initially require any segregation analysis and the minimum IBD region surrounding a selected region is expected to be small (<1 cM) unless the selective sweep was completed in a few generations. However, a problem with this approach is, of course, that it does not show which trait or traits the selected locus affects. Furthermore, false positives will occur because non-selected regions might become fixed

ADMIXTURE
The mixing of two genetically differentiated populations.

INTROGRESSION
The transfer of genetic material from one population to another by repeated backcrossing.

GENETICAL GENOMICS
Segregation analysis of gene expression data using genetic markers to trace the inheritance of individual chromosome segments.

owing to random genetic drift. Therefore, it might be necessary to combine this approach with segregation analysis to confirm the presence of QTLs and to establish genotype–phenotype relationships.

The generation of draft genome sequences will also make it possible to construct comprehensive oligonucleotide arrays for expression analysis. This might turn out to be a useful complement to QTL mapping for identifying genes that underlie complex traits. For example, the threefold upregulation of *IGF2* expression in pigs that carry the intron 3 G3073A mutation[7] indicates that *IGF2* could have been detected as a differentially expressed gene by an array analysis of skeletal-muscle mRNA from the wild boar and domestic pigs. However, a problem with this type of association analysis is that it is not possible to resolve whether an observed differential expression is a direct effect that is due to a *cis*-acting regulatory element, or a secondary effect that is due to changes in gene expression at other loci. More recently this issue has been addressed by an approach that is known as GENETICAL GENOMICS, in which the expression of each individual gene is treated as a trait in QTL analyis[50,51]. The approach is costly as it requires expression analysis of numerous individuals (>100), and there is also a problem in defining the appropriate statistical significance thresholds owing to the many different tests that need to be carried out. Nonetheless, the approach has the potential to uncover interesting *cis*- and *trans*-acting effects of regulatory mutations. Finally, the emerging field of proteomics might provide further tools for the dissection of the molecular basis for phenotypic variation[52].

Conclusions

The generation of complete genome sequences for domestic animals is justified not only by the agricultural importance of these species, but also by the potential contributions of these genome projects to basic biology and human medicine. Genome research in these species is generally regarded as applied science, the aim being to generate agricultural applications. These agricultural applications are certainly the greatest impetus for ongoing genome sequencing programmes in domestic animals, and molecular information will be increasingly important in practical breeding programmes. However, we argue that domestic animals will also contribute to basic biology as they provide unique opportunities for unravelling the genetic basis of phenotypic variation. Furthermore, human medicine will also benefit from the progress in domestic-animal genomics. Domestic animals provide models for disorders with both a monogenic and multifactorial background (see TABLE 1 and the QTLs discussed above). So, post-genomic studies of domestic animals will lead to new knowledge of gene function and biochemical pathways that are altered in disease conditions, and will expand the availability of large-animal models for the testing of new disease treatments.

1. Darwin, C. *On the Origins of Species by Means of Natural Selection or The Preservation of Favoured Races in the Struggle for Life* (John Murray, London, 1859).
2. Darwin, C. *The Variation of Animals and Plants Under Domestication* (John Murray, London, 1868).
3. Darvasi, A. & Pisante-Shalom, A. Complexities in the genetic dissection of quantitative trait loci. *Trends Genet.* **18**, 489–491 (2002).
4. Glazier, A. M., Nadeau, J. H. & Aitman, T. J. Finding genes that underlie complex traits. *Science* **298**, 2345–2349 (2002).
5. Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422**, 835–847 (2003).
6. Grisart, B. *et al.* Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Res.* **12**, 222–231 (2002).
7. Van Laere, A. S. *et al.* Positional identification of a regulatory mutation in *IGF2* causing a major QTL effect on muscle growth in the pig. *Nature* **425**, 832–836 (2003).
8. Blott, S. *et al.* Molecular dissection of a QTL: a phenylalanine to tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* **163**, 253–266 (2002).
9. Grisart, B. *et al.* Genetic and functional demonstration of the causality of the DGAT1 K232A mutation in the determinism of the BTA14 QTL affecting milk yield and composition. *Proc. Natl Acad. Sci. USA* (in the press).
10. Winter, A. *et al.* Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA: diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proc. Natl Acad. Sci. USA* **99**, 9300–9305 (2002).
References 6, 9 and 10 show, by linkage and linkage-disequilibrium analysis combined with functional analysis, that a missense mutation underlies an important QTL for milk yield and composition in cattle.
11. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer associates Inc., Sunderland, 1998).
12. Mackay, T. F. C. Quantitative trait loci in *Drosophila*. *Nature Rev. Genet.* **2**, 11–21 (2001).
13. Andersson, L. Genetic dissection of phenotypic diversity in farm animals. *Nature Rev. Genet.* **2**, 130–138 (2001).
14. Jansen, R. C. in *Handbook of Statistical Genetics*. 2nd edn (eds. Balding, D. J., Bishop, M. & Cannings, C.) 445–476 (Wiley, England, 2003).
15. Hoeschele, I. in *Handbook of Statistical Genetics*. 2nd edn (eds. Balding, D. J., Bishop, M. & Cannings, C.) 477–525 (Wiley, England, 2003).
16. Brockmann, G. A. *et al.* Single QTL effects, epistasis, and pleiotropy account for two-thirds of the phenotypic F(2) variance of growth and obesity in DU6i x DBA/2 mice. *Genome Res.* **10**, 1941–1957 (2000).
17. Carlborg, Ö. *et al.* A global search reveals epistatic interaction between QTLs for early growth in the chicken. *Genome Res.* **13** (2003).
A global two-dimensional search that showed that epistatic interaction explains a considerable portion of the genetic variance for early growth in an intercross between the red junglefowl and White Leghorn chickens.
18. Shimomura, K. *et al.* Genome-wide epistatic interaction analysis reveals complex genetic determinants of circadian behavior in mice. *Genome Res.* **11**, 959–980 (2001).
19. Steinmetz, L. M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–330 (2002).
20. Nezer, C. *et al.* Haplotype sharing refines the location of an imprinted QTL with major effect on muscle mass to a 250 Kb chromosome segment containing the porcine *IGF2* gene. *Genetics* **165**, 277–285 (2003).
21. Fujii, J. *et al.* Identification of a mutation in the porcine ryanodine receptor that is associated with malignant hyperthermia. *Science* **253**, 448–451 (1991).
22. Marklund, S. *et al.* Molecular basis for the dominant white phenotype in the domestic pig. *Genome Res.* **8**, 826–833 (1998).
23. Seitz, J. J., Schmutz, S. M., Thue, T. D. & Buchanan, F. C. A missense mutation in the bovine *MGF* gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. *Mamm. Genome* **10**, 710–712 (1999).
24. Milan, D. *et al.* A mutation in *PRKAG3* associated with excess glycogen content in pig skeletal muscle. *Science* **288**, 1248–1251 (2000).
25. Ciobanu, D. *et al.* Evidence for new alleles in the protein kinase adenosine monophosphate-activated γ3-subunit gene associated with low glycogen content in pig skeletal muscle and improved meat quality. *Genetics* **159**, 1151–1162 (2001).
26. Lindahl, G. *et al.* A second mutant allele (V199I) at the PRKAG3 (RN) locus- I. Effect on technological meat quality of pork loin. *Meat Science* **66**, 609–619 (2003).
27. Lindahl, G. *et al.* A second mutant allele (V199I) at the PRKAG3 (RN) locus- II. Effect on colour characteristics of pork loin. *Meat Science* **66**, 621–627 (2003).
28. Scott, J. W. *et al.* CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J. Clin. Invest.* **113**, 274–284 (2004).
29. Darvasi, A. Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genetics* **18**, 19–24 (1998).
30. Meuwissen, T. H. & Goddard, M. E. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* **155**, 421–430 (2000).
31. Meuwissen, T. H. & Goddard, M. E. Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Sel. Evol.* **33**, 605–634 (2001).
32. Farnir, F. *et al.* Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics* **161**, 275–287 (2002).
33. Perez-Enciso, M. Fine mapping of complex trait genes combining pedigree and linkage disequilibrium information: a Bayesian unified framework. *Genetics* **163**, 1497–1510 (2003).
References 30–33 describe a statistical methodology that exploits linkage disequilibrium for high-resolution mapping in domestic animals.
34. Maynard-Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
A classic paper that shows that a selective sweep of a favourable mutation causes a 'hitch-hiking' effect at linked loci.
35. Meuwissen, T. H., Karlsen, A., Lien, S., Olsaker, I. & Goddard, M. E. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* **161**, 373–379 (2002).

36. Silver, L. M. *Mouse Genetics* (Oxford University Press, New York, 1995).
37. Beck, J. A. *et al.* Genealogies of mouse inbred strains. *Nature Genet.* **24**, 23–25 (2000).
38. Nadeau, J. H. & Frankel, W. N. The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. *Nature Genet.* **25**, 381–384 (2000).
39. Kirkness, E. F. *et al.* The dog genome: survey sequencing and comparative analysis. *Science* **301**, 1898–1903 (2003).
40. Ostrander, E. A. & Kruglyak, L. Unleashing the canine genome. *Genome Res.* **10**, 1271–1274 (2000).
41. Lingaas, F. *et al.* A mutation in the canine *BHD* gene is associated with hereditary multifocal renal cystadeno-carcinoma and nodular dermatofibrosis in the German Shepherd dog. *Hum. Mol. Genet.* **12**, 3043–3053 (2003).
42. Springer, M. S., Murphy, W. J., Eizirik, E. & O'Brien, S. J. Placental mammal diversification and the Cretaceous–Tertiary boundary. *Proc. Natl Acad. Sci. USA* **100**, 1056–1061 (2003).
43. Stone, R. T. *et al.* Use of bovine EST data and human genomic sequences to map 100 gene-specific bovine markers. *Anim. Genet.* **13**, 211–215 (2002).
44. Fahrenkrug, S. C., Freking, B. A., Smith, T. P., Rohrer, G. A. & Keele, J. W. Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. *Anim. Genet.* **33**, 186–195 (2002).
45. Wade, C. M. *et al.* The mosaic structure of variation in the laboratory mouse genome. *Nature* **420**, 574–578 (2002).
46. The International HapMap Consortium. The international HapMap project. *Nature* **426**, 789–796 (2003).
47. Farnir, F. *et al.* Extensive genome-wide linkage disequilibrium in cattle. *Genome Res.* **10**, 220–227 (2000).
48. Giuffra, E. *et al.* The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**, 1785–1791 (2000).
49. Bruford, M. W., Bradley, D. G. & Luikart, G. DNA markers reveal the complexity of livestock domestication. *Nature Rev. Genet.* **4**, 900–910 (2003).
A review on the complexity of livestock domestication.
50. Jansen, R. C. & Nap, J. P. Genetical genomics: the added value from segregation. *Trends Genet.* **17**, 388–391 (2001).
51. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
52. Sellers, T. A. & Yates, J. R. Review of proteomics with applications to genetic epidemiology. *Genet. Epidemiol.* **24**, 83–98 (2003).
53. Olson, M. V. & Varki, A. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nature Rev. Genet.* **4**, 20–28 (2003).
54. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
55. Clark, A. G. *et al.* Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**, 1960–1963 (2003).
56. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
57. Thomas, J. W. *et al.* Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788–793 (2003).
58. Margulies, E. H., Blanchette, M., Haussler, D. & Green, E. D. Identification and characterization of multi-species conserved sequences. *Genome Res.* **13**, 2507–2518 (2003).
References 57–58 show the power of multi-species comparative sequence analysis for detecting evolutionary conserved sequences.
59. Silvers, W. K. *The Coat Colors of Mice. A Model for Mammalian Gene Action and Interaction.* (Springer-Verlag, New York, 1979).
60. Hartman, J. L. T., Garvik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291**, 1001–1004 (2001).
61. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33** (Suppl.), 245–254 (2003).
62. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the *agouti* locus in the mouse. *Nature Genet.* **23**, 314–318 (1999).
63. Grewal, S. I. & Klar, A. J. Chromosomal inheritance of epigenetic states in fission yeast during mitosis and meiosis. *Cell* **86**, 95–101 (1996).
64. Cavalli, G. & Paro, R. The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* **93**, 505–518 (1998).
65. Sollars, V. *et al.* Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nature Genet.* **33**, 70–74 (2003).
66. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**, 157–161 (1999).
67. Kakutani, T. Epi-alleles in plants: inheritance of epigenetic information over generations. *Plant Cell Physiol.* **43**, 1106–1111 (2002).
68. Martienssen, R. A. & Colot, V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**, 1070–1074 (2001).
References 62–68 illustrate the fact that epimutations can be inherited and can contribute to phenotypic variation.
69. Jeon, J.-T. *et al.* A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the *IGF2* locus. *Nature Genet.* **21**, 157–158 (1999).
70. Nezer, C. *et al.* An imprinted QTL with major effect on muscle mass and fat deposition maps to the *IGF2* locus in pigs. *Nature Genet.* **21**, 155–156 (1999).
71. Kijas, J. M. H. & Andersson, L. A phylogenetic study of the origin of the domestic pig estimated from the near complete mtDNA genome. *J. Mol. Evol.* **52**, 302–308 (2001).
72. Grobet, L. *et al.* A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nature Genet.* **17**, 71–74 (1997).
73. Grobet, L. *et al.* Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle. *Mamm. Genome* **9**, 210–213 (1998).
74. Kambadur, R., Sharma, M., Smith, T. P. & Bass, J. J. Mutations in myostatin (*GDF8*) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Res.* **7**, 910–916 (1997).
75. McPherron, A. C. & Lee, S. J. Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl Acad. Sci. USA* **94**, 12457–12461 (1997).
76. Klungland, H., Vage, D. I., Gomez-Raya, L., Adalsteinsson, S. & Lien, S. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mamm. Genome* **6**, 636–639 (1995).
77. Lundén, A., Marklund, S., Gustafsson, V. & Andersson, L. A nonsense mutation in the FMO3 gene underlies fishy off-flavor in cow's milk. *Genome Res.* **12**, 1885–1888 (2003).
78. Tobita-Teramoto, T. *et al.* Autosomal albino chicken mutation (c(a)/c(a)) deletes hexanucleotide (-Delta GACTGG817) at a copper-binding site of the tyrosinase gene. *Poult. Sci.* **79**, 46–50 (2000).
79. Kerje, S., Lind, J., Schütz, K., Jensen, P. & Andersson, L. Melanocortin 1-receptor (MC1R) mutations are associated with plumage colour in chicken. *Animal Genetics* **34**, 241–248 (2003).
80. Kerje, S. *Mapping Genes Affecting Phenotypic Traits in Chicken*. Thesis, Uppsala Univ., Sweden (2003).
81. Lin, L. *et al.* The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* **98**, 365–376 (1999).
82. Schmutz, S. M., Berryere, T. G., Ellinwood, N. M., Kerns, J. A. & Barsh, G. S. MC1R studies in dogs with melanistic mask or brindle patterns. *J. Hered.* **94**, 69–73 (2003).
83. Pailhoux, E. *et al.* A 11.7-kb deletion triggers intersexuality and polledness in goats. *Nature Genet.* **29**, 453–458 (2001).
In this paper, positional cloning showed that an 11.7-kb deletion of an intergenic region causes lack of horns and intersexuality in goats.
84. Marklund, L., Moller, M. J., Sandberg, K. & Andersson, L. A missense mutation in the gene for melanocyte-stimulating hormone receptor (*MC1R*) is associated with the chestnut coat color in horses. *Mamm. Genome* **7**, 895–899 (1996).
85. Rieder, S., Taourit, S., Mariat, D., Langlois, B. & Guerin, G. Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (*Equus caballus*). *Mamm. Genome* **12**, 450–455 (2001).
86. Mariat, D., Taourit, S. & Guerin, G. A mutation in the *MATP* gene causes the cream coat colour in the horse. *Genet. Sel. Evol.* **35**, 119–133 (2003).
87. Santschi, E. M. *et al.* Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. *Mamm. Genome* **9**, 306–309 (1998).
88. Metallinos, D. L., Bowling, A. T. & Rine, J. A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung disease. *Mamm. Genome* **9**, 426–431 (1998).
89. Yang, G. C. *et al.* A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome (LWFS); a horse variant of Hirschsprung disease. *Hum. Mol. Genet.* **7**, 1047–1052 (1998).
90. Johansson Moller, M. *et al.* Pigs with the dominant white coat color phenotype carry a duplication of the *KIT* gene encoding the mast/stem cell growth factor receptor. *Mamm. Genome* **7**, 822–830 (1996).
91. Giuffra, E. *et al.* A large duplication associated with Dominant White color in pigs originated by homologous recombination between LINE elements flanking KIT. *Mamm. Genome* **13**, 569–577 (2002).
92. Hasler-Rapacz, J. *et al.* Identification of a mutation in the low density lipoprotein receptor gene associated with recessive familial hypercholesterolemia in swine. *Am. J. Med. Genet.* **76**, 379–386 (1998).
93. Kijas, J. M. H. *et al.* Melanocortin receptor 1 (*MC1R*) mutations and coat color in pigs. *Genetics* **150**, 1177–1185 (1998).
94. Kijas, J. M. H., Moller, M., Plastow, G. & Andersson, L. A frameshift mutation in *MC1R* and a high frequency of somatic reversions cause black spotting in pigs. *Genetics* **158**, 779–785 (2001).
95. Meijerink, E. *et al.* A DNA polymorphism influencing alpha(1,2)fucosyltransferase activity of the pig FUT1 enzyme determines susceptibility of small intestinal epithelium to *Escherichia coli* F18 adhesion. *Immunogenet.* **52**, 129–136 (2000).
96. Galloway, S. M. *et al.* Mutations in an oocyte-derived growth factor gene (*BMP15*) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature Genet.* **25**, 279–283 (2000).
97. Mulsant, P. *et al.* Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Merino ewes. *Proc. Natl Acad. Sci. USA* **98**, 5104–5109 (2001).
98. Freking, B. A. *et al.* Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.* **12**, 1496–1506 (2002).
99. Charlier, C. *et al.* The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status. *Nature Genet.* **27**, 367–369 (2002).
References 98–99 show that a single nucleotide substitution in an intergenic region upregulates several imprinted genes in cis, which in turn leads to muscle hypertrophy in sheep.
100. Meuwissen, T. H. E & Goddard, M. E. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multiple traits. *Genet. Sel. Evol.* (in the press).

Acknowledgements

We thank K. Lindblad-Toh and two anonymous reviewers for their useful comments, and M. Fredholm, B. Liu, W. Warren and G. Weinstock for their updates on the status of genome sequencing projects.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez: http://www.ncbi.nlm.nih.gov/Entrez
DGAT

LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink
Agouti | *GHR* | *IGF2* | *KITLG* | myostatin | *PRKAG3* | *RYR1*

FURTHER INFORMATION

Leif Andersson's laboratory:
http://www.imbim.uu.se/forskning/anderssonresearch.html
ArkDB: http://www.thearkdb.org
A BAC fingerprint map of the bovine genome:
http://www.bcgsc.ca/lab/mapping/bovine
BBSRC ChickEST Database: http://www.chick.umist.ac.uk
Beijing Genomics Institute:
http://www.genomics.org.cn/bgi/engfish2/index.htm
Bovine genome project:
http://www.hgsc.bcm.tmc.edu/projects/bovine
Bovine genome project:
http://bos.cvm.tamu.edu/bovgbase.html
Broad Institute dog genome project:
http://www.broad.mit.edu/media/2003/pr_03_tasha.html
Cat genome project:
http://rex.nci.nih.gov/lgd/cat/catgenome.htm
Chicken genome project:
http://genome.wustl.edu/projects/chicken
ChickNET: http://www.chicken-genome.org/resources
CSIRO cattle genomics:
http://www.livestockgenomics.csiro.au/cattle.shtml
Dog genome project: http://mendel.berkeley.edu/dog.html
DogMap: http://www.dogmap.ch
GenBank's trace archive: http://www.ncbi.nlm.nih.gov/Traces
Horse genome project: http://www.uky.edu/Ag/Horsemap
Mapping the caprine genome:
http://dga.jouy.inra.fr/cgi-bin/lgbc/intro2.pl?BASE=goat
NHGRI genome sequencing proposals:
http://www.genome.gov/10002154
USDA MARC gene mapping information:
http://www.marc.usda.gov/genome/genome.html
UK sheep genome mapping project:
http://www.projects.roslin.ac.uk/sheepmap/front.html
US pig gene mapping coordination program:
http://www.genome.iastate.edu/pig
US poultry genome project:
http://poultry.mph.msu.edu
The Institute for Genomic Research: http://www.tigr.org
Access to this interactive links box is free online.